Automation and Remote Control, Vol. 55, No. 7, 1994

NONLOCAL-OPTIMIZATION METHODS OF THE FIRST ORDER BASED ON POTENTIAL THEORY

A. I. Kaplinskii and A. I. Propoi

UDC 517.977.5

Nonlocal-optimization methods based on the variation of the first order of the functional of a randomized problem are considered. Gradient descent methods with respect to a potential function (regarded as a model of extremal properties of the objective function), "tunneling" algorithms, and an adaptive generalization of the variable polyhedron method are studied. All these methods require measurements of just values of the objective function. Structural characteristics of nonlocal search methods of the first order are analyzed based on the properties of the potential function. This work is a continuation of [1].

1. INTRODUCTION

We consider the problem

$$f(x) \to \min_{x \in R^n} \tag{1}$$

and its randomization

$$F(X) = E[f(X)] \to \min_X,$$

which is equivalent to (1) [1]. Here, X is a random vector with values in R^n and with the probability distribution function $p(x)$, and E is the symbol of mathematical expectation. The first-order variation $X_\varepsilon = X + \varepsilon Y$ of the functional F gives the condition of improvement [1]:

$$\delta_Y F(X) = -\int_{R^n} (f(x) - c)\operatorname{div}[p(x)\bar{y}(x)]\,dx \le 0, \tag{2}$$

where the existence of the joint distribution function $p(x, y) = p(x)p(y|x)$ is assumed, and

$$\bar{y}(x) = \int_{R^n} yp(y|x)\,dy. \tag{3}$$

$$\int_{R^n} \operatorname{div}[p(x)\bar{y}(x)]\,dx = 0. \tag{4}$$

To satisfy (2) we set

$$\operatorname{div}[p(x)\bar{y}(x)] = \Phi(f(x) - c)p_U(x), \tag{5}$$

Voronezh State University. Institute for System Analysis, Russian Academy of Science, Moscow. Translated from Avtomatika i Telemekhanika, No. 7, pp. 94–103, July, 1994. Original article submitted November 22, 1993.

0005-1179/94/5507-1004$12.50

where the function $p_U(x) \geq 0$ and can be interpreted as the probability distribution function of a random vector U; $\Phi(t)$ is a function of the scalar argument t with properties

$$\Phi(0) = 0, \quad t\Phi(t) \geq 0. \tag{6}$$

We choose functions Φ and p_U that ensure the existence of the integral $\int_{R^n} (f(x) - c)\Phi(f(x) - c)p_U(x)\,dx$ for the given function f and finite values of c. The condition of improvement

$$\int_{R^n} (f(x) - c)\Phi(f(x) - c)p_U(x)\,dx \geq 0$$

is obviously fulfilled for these functions due to (6).

In order to satisfy (4) it is necessary that

$$\int_{R^n} \Phi(f(x) - c)p_U(x)\,dx = Ex[\Phi(f(U)) - c] = 0. \tag{7}$$

The above relations give a basis for developing first-order numerical methods of nonlocal optimization.

2. IMPLEMENTATION OF METHODS OF NONLOCAL SEARCH

We shall consider nonlocal search methods within the framework of probabilistic iterative procedures. Let k be the number of the iteration. For the known random vector X_k (the state of the search system) we choose a random vector U_k (a control) that statistically depends on the vector X_k in the general case:

$$p_{X_k,U_k}(x, u) = p_{X_k}(x)p_{U_k|X_k=x}(u).$$

For a fixed distribution of the random vector U_k the relation (7) is a nonlinear regression equation with respect to the level c. We denote the current estimate of its solution by c_k.

The condition of improvement (5) for $c = c_k$, $U = U_k$ or its variants [7] make it possible to construct a vector field $\bar{y}_k(x) = \bar{y}_k(x; c_k)$ that determines the dynamics of the random vectors X_k:

$$X_{k+1} = X_k + \varepsilon_k \bar{y}_k(X_k). \tag{8}$$

In this case $F(X_{k+1}) \leq F(X_k)$ for the corresponding choice of the step length ε_k.

We note that the motion (8) is determined formally so far because it requires the specification of characteristics in which the random vector X_k is measured (and controlled).

The motion (8) implies the transformation of the distribution function of X_k given by the equation

$$p_{X_{k+1}}(x) = p_{X_k}(x) - \varepsilon \mathrm{div}[p_{X_k}(x)\bar{y}_k(x)]. \tag{9}$$

This equation is a consequence of the first-order variation of the density function $p_{X_k+\varepsilon\bar{y}_k(X_k)}(x) = \int_{R^n} p(x - \varepsilon y, y)\,dy$ of the random vector $X_{k+1} = X_k + \varepsilon\bar{y}_k(X_k)$: (it can be also considered as a discrete approximation with respect to "time" ε of the continuity equation [3]).

The relations (5), (7), and (9) define a closed system of equations (with respect to p_{X_k}, p_{U_k}, and c_k) that specify the organization of nonlocal search.

We will consider the case where the potential component is singled out in the vector field $\bar{y}(x)$ [1]:

$$p(x)\bar{y}(x) = \nabla\varphi(x) + w(x), \qquad \mathrm{div}\ w(x) = 0, \tag{10}$$

where $\varphi(x)$ is the solution of the Poisson equation: $\Delta\varphi(x) = \Phi(f(x) - c)p_U(x)$, Δ is the Laplace operator. This solution can be represented in integral form (for $\varphi|_\infty = 0$) as

$$\varphi(x) = \int_{R^n} G(x,\xi)\Phi(f(\xi) - c)p_U(\xi)\,d\xi = E[G(x,U)\Phi(f(U) - c)], \tag{11}$$

where for $n > 2$

$$G(x,\xi) = -\frac{1}{(n-2)\omega_n\|x-\xi\|^{n-2}} \tag{12}$$

is the fundamental solution of the Laplace equation [2].

Remarks. 1. Other ways of fulfilling of the condition of improvement are that lead to other types of potential function (in particular, to the wave potential) are also possible [1, 3].

2. In the general case the potential function is determined by the solution of a boundary-value problem and its integral representation is given by the corresponding Green's function [1–3].

It was shown in [1, 3] that the potential function φ determined by (10)–(12) or, in the general case, by the solution of a boundary-value problem serves as a model of extremal properties of the objective function f. Therefore, analysis of relations (5), (7), and (9) (together with (10)–(12)) allow us to reveal structural components that must be found in nonlocal search procedures [4]. These components make it possible to specify modules that serve as a basis for developing nonlocal search algorithms [5]. These modules can be used simultaneously (forming more complex modules) or sequentially, at separate stages of the nonlocal search process. In any case it is necessary to specify "stationary" states of the search that allow us to judge the termination of a module operation and the necessity of turning to another.

In the present work we consider, on the basis of this approach, the gradient descent method (with respect to the potential function), tunneling algorithms, and an adaptive generalization of the variable polyhedron method.

3. GRADIENT ALGORITHMS

Gradient algorithms can be written formally as

$$X_{k+1} = X_k + \varepsilon_k \nabla\varphi_k(X_k) \tag{13}$$

(this follows from Eqs. (8) and (10)), where in the case of the potential defined by (11) we have [1, 2]

$$\nabla\varphi_k(x) = \frac{1}{\omega^n}\int_{\Omega_k} \frac{\Phi(f(\xi) - c)}{\|x-\xi\|^n}(x-\xi)p_{U_k}(\xi)\,d\xi, \tag{14}$$

or

$$\nabla\varphi_k(x) = \frac{1}{\omega^n} E\left[X_{\Omega_k}(U_k)\Phi(f(U_k) - c)\frac{x - U_k}{\|x-U_k\|^n}\right]. \tag{15}$$

Here, $\nabla\varphi$ is the gradient of the function φ, X_{Ω_k} is the indicator (the characteristic function) of the set Ω_k, and

$$\int_{\Omega_k} p_{U_k}(\xi)\,d\xi = 1.$$

The natural way to implement (13)–(15) is to use probabilistic iterative methods (of the stochastic approximation type) [5–7]. In this case the representation of $\nabla\varphi$ as the mathematical expectation (15) makes it possible not to compute the integral in (13) but to use samples of the random vector U_k (the distribution of the vector U_k must be fixed at this stage).

The algorithm consists of the following steps.

1. Choose an initial distribution of the random vector U^0. In this case (7) is a nonlinear regression equation with respect to the level c. Denote the current estimate of the level c by c^0. Methods of solving these equations are well known [9]. In these methods the function Φ serves as a controlled parameter ensuring the robustness of the algorithm [9].

2. For the given initial state x^0 (a realization of the random vector X^0) and the chosen distribution of U^0, compute the subsequent values x_k $(k = 1, 2, 3, \ldots)$ from (13) and (14). For the case where $\Omega_k = R^n$ we have

$$x_{k+1} = x_k + \varepsilon_k \Phi(f(u_k^0) - c^0) \frac{x_k - u_k^0}{\omega_n \|x_k - u_k^0\|^n}, \tag{16}$$

where u_k^0 is a realization of the random vector U_k^0 at the k-th iteration. We note that in (16) the gradient of the fixed potential function $\nabla\varphi_0(x) = \nabla\varphi(x; c^0)$ is used.

The level c divides the search space R^n into two, possibly multiply connected, domains: the perspective domain $\Omega_P(c)$ and the nonperspective domain $\Omega_N(c)$:

$$\begin{aligned} \Omega_P(c) &= \{x | f(x) \le c\}, \\ \Omega_N(c) &= \{x | f(x) > c\}. \end{aligned} \tag{17}$$

It can be shown, by using results on the convergence of probabilistic iterative procedures [6–9], that the initial distribution $p_{X^0}(x)$ is transformed by the process (16), (9) into the distribution $p_{X^1}(x)$ concentrated in the perspective domain $\Omega_P(c^0)$ in one or another probabilistic sense. The convergence analysis of the algorithm (16) was done in [10] and partially in [4]. We note that in this case the function Φ is used for constructing optimal probabilistic algorithms [8, 9].

After termination of this stage of nonlocal search we may change the potential function: $\varphi = \varphi(x, c^1)$, setting, e.g., $U^1 = X^1$ and estimating c^1 for this distribution. For continuation of the process it is necessary to examine the structure of the potential function.

4. STRUCTURAL COMPONENTS OF THE FIRST-ORDER METHODS

We partition the function φ into two components

$$\varphi(x) = \int_{R^n} G(x, \xi) \tilde{f}(\xi) p_U(\xi)\, d\xi = \varphi_P(x) + \varphi_N(x), \tag{18}$$

where

$$\varphi_P(x) = \int_{\Omega_P} G(x, \xi) \tilde{f}(\xi) p_U(\xi)\, d\xi, \tag{19}$$

$$\varphi_N(x) = \int_{\Omega_N} G(x, \xi) \tilde{f}(\xi) p_U(\xi)\, d\xi, \tag{20}$$

$\tilde{f}(\xi) = \Phi(f(\xi) - c)$, and the sets Ω_P and Ω_N are defined in (17).

From the definition of these sets and properties of the solution of the Poisson equation (11) it follows that $\varphi_N(x) \le 0$ and $\varphi_P(x) \ge 0$ for all $x \in R^n$. From the properties of the Newton potentials (19), (20) we obtain that

$$\Delta\varphi_N(x) \begin{cases} \ge 0, & x \in \Omega_N, \\ = 0, & x \in \Omega_P. \end{cases} \qquad \Delta\varphi_P(x) \begin{cases} = 0, & x \in \Omega_N, \\ \le 0, & x \in \Omega_P, \end{cases}$$

Thus, φ_N is subharmonic in R^n and harmonic in Ω_P; φ_P is superharmonic in R^n and harmonic in Ω_N [11].

According to the maximum principle for subharmonic functions [11], the maximum value of $\varphi_N(x)$ is not achieved inside the domain Ω_N, i.e., search motions along the direction $\nabla\varphi_N(x)$ lead to expulsion of the search point onto the boundary of nonperspective and perspective sets. In this case

$$\sup \varphi_N(x) = 0 \tag{21}$$

(since $\tilde{f}(x) > 0$ in Ω_N and $G(x, \xi) p_U(\xi) < 0$ in the entire space R^n). Here, x is a realization of the random vector X, obtained, e.g., by iteration of (16) and, consequently, concentrated in the perspective domain Ω_P.

The potential in (21) is defined by the level c and the random vector U. Therefore, the problem is to find a random vector U, statistically connected with X in the general case, that satisfies the equality (21) for the fixed level c, i.e.,

$$\int_{\Omega_N} \Phi(f(\xi) - c) G(x, \xi) p_U(\xi)\, d\xi = 0. \tag{22}$$

Introducing the indicator $\chi_{\bar{\Omega}_N}(x)$ of the closure $\bar{\Omega}_N$ of Ω_N we rewrite (22) as

$$E\left[\chi_{\bar{\Omega}_N}(U)\Phi(f(U)-c)G(x,U)\right]=0. \tag{23}$$

According to (21) and (23) determination of the distribution of U can be done in two ways.

First, if we parameterize the density function of the random vector U, then (23) will be a nonlinear regression equation with respect to this parameter and therefore it can be solved by the same iterative methods that are used for solution of Eq. (7).

The second way is to use gradient descent of the type (16), but, according to (21), for the component of the potential function $\varphi_N(x)$ defined in (20).

In any case, according to the maximum principle for subharmonic functions, the search procedure must lead the search point to the boundary of the perspective and nonperspective domains (for the given level c and the fixed point x).

The solution of Eqs. (21) and (23) must be accomplished together with the solution of Eq. (7) that defines the level c. Because these solution processes are separated, we rewrite (7) as

$$E\left[\Phi(f(V)-c)\right]=0, \tag{24}$$

i.e., in (24) we define the level c for the given distribution $p_V(\xi)$ of the random vector V, whereas in (23) for the given level c we determine the distribution $p_U(\xi)$ of the random vector U that "zeroes" $\varphi_N(x,U)$ in accordance with (21).

In particular, if $\Phi(t)=t$ in (24), then $c=E[f(V)]$ is the average value of the function f with respect to the random vector V. In this case we obtain from (23), (24)

$$E\left[\chi_{\bar{\Omega}(E[f(V)])}(U)G(x,U)(f(U)-f(V))\right]=0. \tag{25}$$

If the solution of (25) is sought in the class of δ-functions, i.e., $p_U(\xi)=\delta(\xi-u)$, then it is necessary to solve the nonlinear deterministic equation

$$G(x,u)\Phi(f(u)-c)=0. \tag{26}$$

Let x be the point of a local minimum of f. Then, in view of properties of the kernel $G(x,u)$ (having a singularity at x) and the function Φ, the solution of (26) will be a point $u\neq x$ for which $f(u)=c$. This transition $x\to u$ is called the tunneling phase [12–14]. The solution of Eq. (23) or the joint solution of Eqs. (23) and (24) (or Eq. (25)) provides a probabilistic extension of the tunneling procedure, realizing the transition $X\to U$ with conservation of the mean value of the function f in the case of (25).

In particular, if in (23) the distribution of the random vector U is delta-shaped: $\frac{1}{\beta^n}p_U((\xi-u)/\beta)\to\delta(\xi-u)$, $\beta\to 0$, then the solution of (23) gives the distribution of U for which $\bar{u}=\int_{R^n}\xi p_U(\xi)\,d\xi\neq x$, but $\int_{R^n}f(\xi)p_U(\xi)\,d\xi=c$.

Remarks. 1. As we noted above, the solution of Eq. (26) that realizes the transition $x\to u$ is singular. For its regularization it is desirable to preserve the integral representation (23) realizing the transition $X\to U$. This makes it possible to pass from the random variable $G(x,U)\Phi(f(U)-c)$, having generally an infinite variance (because of the singularity in $G(x,U)$), to a new random value with the same mathematical expectation but with finite variance. This is implemented by inclusion of the singularity into the density function [5].

2. In the general case the power of the denominator in the kernel $G(x,\xi)$ is used as a parameter:

$$\nabla_x G(x,\xi)=\frac{x-\xi}{\omega_n\|x-\xi\|^\lambda},\qquad 0\le\lambda\le n$$

(cf. (12)), i.e., the parameter λ defines the "real" dimension of the search manifold.

3. As was shown in [1, 3], it is possible to use different potential functions (in particular, the wave potential). Such potential functions were used on a heuristic basis in tunneling algorithms [12–14].

4. The successive application of the tunneling phase leads to the transformation of the objective function with the kernel $\prod_i G(x_i,u)$, where x_i are samples of stationary states X^i (associated, in particular, with local minima of f).

Therefore the implementation of the nonlocal search is subdivided into two main phases: the phase of localization of solution (i.e., localization of a subset in $\Omega_P(c)\cap\bar{\Omega}_N(c)$ or a distribution of X on this subset) and the nonlocal phase ensuring the expulsion of the search point out of this subset with conservation of the mean value of the objective function in the case where $\Phi(t)=0$. In particular, such a subset can be the vicinity of a local minimum of f.

5. ADAPTIVE GENERALIZATION OF THE VARIABLE POLYHEDRON METHOD

In the algorithm (16) we consider the trajectories $\{x_k\}$ of a nonlocal search process generated by a single realization x^0 of the random vector X^0 (and randomized by the random vector U). It is natural to also consider methods that use a set (beam) of trajectories generated by a set of realizations $\{x_i^0\}$, i.e., to pass to the general case (13). Since the motion (13) is an unobservable process, in this case it is necessary to use estimates of moments of the corresponding distribution functions.

Going over to mathematical expectations in (8), we obtain

$$m_{k+1} = m_k + \varepsilon_k \hat{E}_{X_k}[\hat{y}_k(X_k)], \tag{27}$$

where $m_k = E[X_k]$, m_{k+1}, $\hat{E}_{X_k}$ are statistical estimates of the corresponding random values.

Using, as before, the gradient of a potential function for determination $\hat{y}_k(x)$, we have

$$m_{k+1} = m_k + \varepsilon_k \hat{E}_{X_k} E_{U_k}\left[(f(U_k) - c)\frac{X_k - U_k}{\omega_n \|X_k - U_k\|^n}\right]. \tag{28}$$

If we limit ourselves to separate realizations of the random vector U_k and choose, as a realization of X_k, its mathematical expectation, then (28) is written as (cf. (16))

$$m_{k+1} = m_k + \varepsilon_k \frac{f(u_k) - c}{\omega_n \|m_k - u_k\|^n}(m_k - u_k). \tag{29}$$

In accordance with (19) and (5) the process (29) must be constructed in such a way that the distribution of the random vector X_{k+1} is concentrated within the perspective domain $\Omega_P(c)$. Therefore in (29) for the evaluation of m_{k+1} only successful (perspective) motions from the point m_k must be used. The degree of perspectiveness is defined by the deviation of the value of the objective function from its (mean) level c, i.e., by the value of $f(U_k) - c$.

The above reasoning serves as a basis for the adaptive generalization of the variable polyhedron method. The scheme suggested consist of the following stages.

1. Obtain K realizations $\{u_k^i\}$, $i = 1, 2, \ldots, K$, of the random vector U_k, $k = 0, 1, 2, \ldots$. For $K = n+1$ these realizations can be interpreted as vertices of a variable polyhedron in R^n.

2. Compute values of the objective function at these points: $f(u_k^i)$, $i = 1, 2, \ldots, K$, and enumerate them in ascending order of the values of the objective function: $f(u_k^K) \geq f(u_k^{K-1}) \geq \ldots \geq f(u_k^1)$.

3. Compute the mean value (over all realizations)

$$c_k = \frac{1}{K}\sum_{i=1}^{K} f(u_k^i). \tag{30}$$

4. Define the set of perspective realizations

$$\mathcal{L}_k = \{i | f(u_k^i) \leq c_k, \quad i = 1, 2, \ldots, K\},$$

and compute the mean value (center of gravity of the polyhedron) over these realizations:

$$m_k = \frac{1}{L_k}\sum_{i=1}^{L_k} u_k^i, \quad \{i, \ldots, L_k\} = \mathcal{L}_k. \tag{31}$$

5. Apply the procedure (29) to each vertex u_k^i, $i = 1, 2, \ldots, K$:

$$m_{k+1}^i = m_k + \varepsilon_k \frac{m_k - u_k^i}{\|m_k - u_k^i\|^n}[f(u_k^i) - c_k] \tag{32}$$

and refine the step length if an unsatisfactory result is obtained.

If the realization u_k^i is perspective: $f(u_k^i) \leq c_k$, then the motion (32) implements a shift of m_k along the direction $-m_k + u_k^i$, which corresponds to the dilatation stage in the variable polyhedron method.

In the opposite case, when $f(u_k^i) > c_k$, the motion (32) implements a shift of m_k in the opposite direction: $m_k - u_k^i$, which corresponds to the reflection stage in the variable polyhedron method.

The operations (30), (31) are further repeated for the points $\{m_{k+1}^i\}$ and the value m_{k+1} is estimated on the basis of the information obtained.

If the point $\dot{m}_{k+1}$ obtained as a result of this motion satisfies the condition $f(\dot{m}_{k+1}) \leq c_k$, then we set $m_{k+1} = \dot{m}_{k+1}$ and the process is continued. If an unsatisfactory estimate is obtained: $f(\dot{m}_{k+1}) > c_k$, then in order to elucidate the perspectiveness of the chosen direction it is necessary to diminish the length of the step ε_k, which corresponds to the contraction of the variable polyhedron.

If the run of these attempts does not lead to a successful value of $\dot{m}_{k+1}$ then it makes sense to use a new realization of U_k.

6. As a termination criterion we can use the square deviation [2]

$$D_k = \frac{1}{K}\sum_{i=1}^{K}(f(u_k^i) - c_k)^2$$

(it is the estimates of the derivative of $F(X)$ along the vector field $\bar{y}_k(x)$ [1, 3]).

Thus, in the variable polyhedron method different strategies of information processing for perspective and nonperspective domains are implemented. These strategies serve as a basis for construction of a set of algorithmic modules. Then, to modify the variable polyhedron method will mean to specify variants of these modules.

We consider several of these variants.

Determination of the level c. The strategy of specifying c plays a central role in implementation of nonlocal search algorithms (because the value of c determines the partition of the search space into nonperspective and perspective domains). The basis for determination of the level c is the regression equation (7), in which the choice of the function Φ allows us to take into account the level of information about the objective function and to ensure the robustness of the algorithm with respect to the choice of c. The method of selecting Φ is developed in the theory of robust estimation [9]. In particular, if $\Phi(t) = \operatorname{sign} t$, then in the problem of nonlocal search we can compare just values of f. In this case equation (7) gives an estimate of the median of the random value f over a sample of its values. The goal of specifying Φ is to provide a compromise between the sensitivity of the algorithm to the strategy of information processing and the robustness of the algorithm.

If it is necessary to implement the strategy of specifying "record" values of c (in order to predict the minimum value of f), then, selecting the characteristic function of the perspective set Ω_P as the function Φ, we obtain from (7)

$$E[\chi_{\Omega_P}(U)(f(U) - c)] = 0.$$

By using an iterative procedure of stochastic approximation type we have

$$c_{k+1} = c_k - \chi[f(u_k) - c_k],$$

where $\chi(t) = 1$ if $t < 0$ and $\chi(t) = 0$ if $t \geq 0$. This procedure of changing c is used in the method of global minimization suggested in [15].

In the process of nonlocal search the dimension of the polyhedron may become less then the dimension of the search space, in particular, when a (local) minimum is localized. In this case it is advisable to leave the vicinity of this minimum and to reform the polyhedron with conservation of the level c. This can be done by implementing the tunneling phase $m_k \to m_{k+1}$ and forming a polyhedron with center at m_{k+1}.

The revisable mathematical expectation m_k, as was mentioned, can be interpreted as a "center of gravity" of perspective realizations. Therefore we can use for specifying m_{k+1} not only a single estimate $\dot{m}_{k+1}$ but the average of several estimates, satisfying the condition $f(\dot{m}_{k+1}) < c_k$ and corresponding to different realizations of U_k. This strategy corresponds to a modification of the variable polyhedron method, in which several vertices of the polyhedron are reflected.

6. CONCLUSION

In the present work we considered separate stages and modules of nonlocal search that followed from the analysis of structural properties of the first-order variation of a randomized functional. A shortcoming of first-order methods is the poor accounting for the geometry of the objective function (the conditionality of its Hessian), which is expressed in the lack of "memory" about perspective and nonperspective directions. The second-order variation allows one to introduce a unified statistical operator of dilatation-reflection of the search space and on this basis to develop methods of nonlocal search that take into consideration the "memory" about perspective and nonperspective directions [4, 5].

REFERENCES

1. A. I. Kaplinskii and A. I. Propoi, "Methods of nonlocal search based on potential theory," *Avtomat. Telemekh.*, No. 7, 55–64 (1993).
2. R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Interscience, NY (1953).
3. A. I. Kaplinskii and A. I. Propoi, *The Variational Approach to Construction of Nonlocal Search Algorithms* [in Russian], VNIISI, Moscow (1986).
4. A. I. Kaplinskii and A. I. Propoi, *Structural Components of Methods for Nonlocal Search Using Potential Theory* [in Russian], VNIISI, Moscow (1990).
5. A. I. Kaplinskii and A. I. Propoi, *Development of Numerical Methods for Nonlocal Search Using Potential Theory* [in Russian], VNIISI, Moscow (1990).
6. Ya. Z. Tsypkin, *Foundations of the Theory of Learning Systems* [in Russian], Nauka, Moscow (1970).
7. M. B. Nevel'son and R. Z. Has'minskii, *Stochastic Approximation and Recurrent Estimation* [in Russian], Nauka, Moscow (1972).
8. B. T. Polyak and Ya. Z. Tsypkin, "Optimal pseudogradient algorithms," *Avtomat. Telemekh.*, No. 8, 74–84 (1980).
9. Ya. Z. Tsypkin, *Foundations of Information Theory of Identification* [in Russian], Nauka, Moscow, 1984.
10. A. I. Kaplinskii, A. M. Pesin, and A. I. Propoi, "On the methods of nonlocal search," in: *Models and Methods of Optimization* [in Russian], VNIISI, Moscow, pp. 35–45 (1991).
11. W. K. Hayman and P. B. Kennedy, *Subharmonic Functions* , Academic Press, NY (1976).
12. A. V. Levy and A. Montalvo, "The tunneling algorithm for the global minimization of functions," SIAM J. Sci. Stat. Comp., **6**, No. 1, 15–29 (1985).
13. R. P. Ge and Y. F. Qin, "A class of filled functions for finding global minimizers of a function of several variables," *J. Optim. Theory Appl.*, **54**, No. 2, 241–252 (1987).
14. A. A. Zhiglyavskii and A. G. Zhilinskas, *Methods of Search for a Global Extremum* [in Russian], Nauka, Moscow (1991).
15. V. K. Chichinadze, *Solution of Nonconvex Nonlinear Problems of Optimization* [in Russian], Nauka, Moscow (1983).